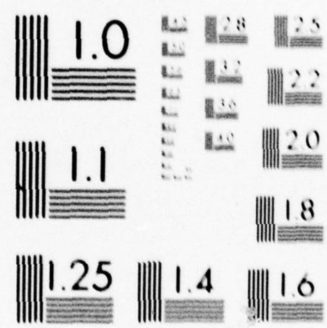
AD-A071 744 PRINCETON UNIV N J DEPT OF STATISTICS F/G 9/2
FORTRAN SUBROUTINES FOR M ESTIMATORS IN THE LINEAR MODEL.(U)
APR 79 A MARRAZI DAAG29-76-G-0298
UNCLASSIFIED TR-147 ARO-14244.7-M NL

1 OF 1
AD
A071744

END
DATE
FILMED
8-79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (19) 14244.7-M (18) ARO	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) (6) Fortran Subroutines for M Estimators in the Linear Model.	5. TYPE OF REPORT & PERIOD COVERED (9) Technical Report,
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Alfio/Marrazi	8. CONTRACT OR GRANT NUMBER(s) (15) DAAG29-76-G-0298
9. PERFORMING ORGANIZATION NAME AND ADDRESS Princeton University Department of Statistics Princeton, New Jersey 08540	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P. O. Box 12211 Research Triangle Park, NC 27709	12. REPORT DATE (11) Apr 1979
	13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 26p.	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

LEVEL II

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(14) TR-147, SER-2

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Fortran subroutines
estimators
linear models
computational methods
least squares problems

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes the computational methods used in programming a set of subroutines to solve linear least squares problems and their robust versions by use of M estimators.

406 873

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

FORTRAN SUBROUTINES FOR M ESTIMATORS IN THE LINEAR MODEL

by

Alfio Marrazi*
Princeton University

Technical Report No. 147, Series 2
Department of Statistics
Princeton University
April 1979

Research was supported in part by a contract with the U. S. Army Research Office, No. DAAG29-76-G-0298, awarded to the Department of Statistics, Princeton University, Princeton, New Jersey, and in part by a grant from the Swiss National Science Foundation, No. 2.569-0.76.

*On leave from the Fachgruppe für Mathematische Statistik, Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland.

1. INTRODUCTION

This report describes the computational methods used in programming a set of subroutines to solve linear least squares problems and their robust versions by use of M estimators. That is, we will concern ourselves with the following problems:

Problem A

Given a real $m \times n$ matrix A of rank $k \le \min(m,n)$, and given a real m-vector $\underline{b}$, find a real m-vector $\hat{\underline{x}}$ minimizing the Euclidean length of $A\underline{x} - \underline{b}$.

(We shall use the symbolism $A\underline{x} \cong \underline{b}$ [1] to denote Problem A.)

Problem B

Solve the following system for $\underline{x} = (x_1, \ldots, x_n)^T$ and σ:

$$\sum_{i=1}^{m} \psi\left(\frac{z_i}{\sigma w_i}\right) a_{ij} w_i = 0 \quad , \quad j = 1 \ldots n \tag{1.1}$$

$$\sum_{i=1}^{m} \chi\left(\frac{z_i}{\sigma w_i}\right) w_i^2 = \text{CONST} \tag{1.2}$$

where

$(a_{ij}) = A$

$z_i = b_i - \sum_{j=1}^{n} a_{ij} x_j$

the w_i's are given weights

$$\psi(x) = \rho'(x) \ ; \ \chi(x) = x\psi(x) - \rho(x) \tag{1.3}$$

CONST is a given number

[1] Usually we shall use, both in the text and the programs, the same notation of L6.

and ρ is a given function, which should satisfy

$$\rho \ge 0 \ , \ \rho \text{ convex} \ , \ \rho(0) = 0$$

$$0 < \lim_{|x| \to \infty} \frac{\rho(x)}{|x|} = L \le \infty \ .$$

When all the w_i-s are equal to 1, the solution $\underline{x}$ of Problem B is an M estimate of the parametervector $\underline{\theta} = (\theta_1 \ldots \theta_n)$ in a linear model $\underline{b} = A\underline{\theta} + \underline{error}$ as defined in L4.

The solution of Problem A may be used as starting value for an iterative algorithm, based on the algorithm H of L4, to solve Problem B.

Problem A is solved by triangularizing A. We use Householder transformations as defined in L6. The basic algorithm for their computation and application is algorithm H12 of that book. By this method we can compute the "minimal length solution" when A is not a full rank matrix.

The introduction of the weights w_i-s in Problem B allows us to compute M estimates of $\underline{\theta}$ which are optimal for a kind of extremal problem proposed in L1: minimize the trace of the asymptotic covariance matrix over all Fisher-consistent (M-) estimators of $\underline{\theta}$ with the same bound on the influence curve considered as a function of $(b,\underline{a})$ ("robustness in factor space"). Here $\underline{a}$ denotes a general row of A, and b the corresponding component of $\underline{b}$. We will briefly discuss how the w_i's should be chosen, but the subroutines for their computation as well as for the computation of the estimate asymptotic covariance matrix have still to be written.

The subroutines are written in FORTRAN following the guidelines in L7 in order to facilitate transportability, readibility and

reliability. Particular attention has been given to modularity for the following reasons:

- Obtain easily several computational schemes (e.g., perform the initial iterations of the algorithm of Problem B using a monotone ψ and the last ones with a redescending ψ, etc.).
- Facilitate further changes in the code.
- Facilitate further additions and particularly the development of a package.

The subroutines are written as parts of a library and belong to three chapters:

1. Main subroutines

MTRF: upper triangularization of A and determination of its pseudorank.

CLLS: solution of Problem A ("Classical Least Squares Solution").

HREG: iterative algorithm for the solution of Problem B.

UCVY: computation of the Unscaled Covariance matrix of the parameter estimates in a regression problem with upper triangular design matrix.

SCVX: computation of the Scaled Covariance matrix of the parameter estimates.

RLLS: examples of some possible combination of MTRF, CLLS, HREG, UCVY and SCVX.

2 and 3. Auxiliary and utility subroutines

These subroutines have very specific purposes and will not be described in this report. Only a brief guide in Appendix 1 indicates their purposes.

The code of MTRF, CLLS and UCVY is very similar to the code

given in L6 for the same purposes. For this reason the description of MTRF, CLLS and UCVY is brief and need only introduce HREG.

The reader who is already familiar with the mentioned statistical and numerical problems can go directly to the documentation on comment cards. This is available on tape together with the programs.

2. SOLVING THE LEAST SQUARES PROBLEM

The computational methods used in the subroutines MTRF and CLLS are those of L6. The following parts of this book contain the theoretical background: Chapters 1, 2, 3, 10 (p.53-57), 11, 13, 14, and part of Chapter 25. We consider the principal steps.

(i) Construction of an $m \times m$ orthogonal matrix Q, of an $n \times n$ permutation matrix P, and of an $m \times n$ matrix R such that

- $QAP = R$.
- R is upper triangular (or upper trapezoidal when $m < n$) .
- the diagonal elements of R are non-increasing in magnitude.

(ii) Determination of the pseudorank of A . This is the largest index j such that the absolute value of the element (j,j) of R is greater than a given value τ . (The appropriate selection of the parameter τ is discussed in L6, ch.25. We will set $\tau = 0$ if the rank of A is complete.)

Let k be the pseudorank of A . Then we write

$$R = \begin{array}{c} \left[\begin{array}{cc} R1 & R2 \\ 0 & R3 \end{array}\right] \begin{array}{l} \}\, k \\ \}\, m-k \end{array} \\ \begin{array}{cc} \underbrace{\quad}_{k} & \underbrace{\quad}_{n-k} \end{array} \end{array} \tag{2.1}$$

with $R1$ upper triangular (k by k)

$R2$ k by $(n-k)$

$R3$ $(m-k)$ by $(n-k)$.

If $k = \min(m,n)$ the matrix $R3$ does not appear in (2.1). If $k = n<m$ only the matrix $R1$ will appear in (2.1). In this case, the computations in (iii) are skipped and we set $W = R1$, $V = I_n$ (here I_n is the n by n identity matrix).

(iii) The choice of a pseudorank $k < \min(m,n)$ means that $R3$ may be ignored. In this case we consider a new least squares problem $A1\underline{x} \cong \underline{b}$ replacing the original $A\underline{x} \cong \underline{b}$, where

$$A1 = Q^T \begin{bmatrix} R1 & R2 \\ 0 & 0 \end{bmatrix} P^T \tag{2.2}$$

is a new design matrix replacing $A = Q^T R P^T$. We compute then an $n \times n$ orthogonal matrix V (denoted by K in L6) and a $k \times k$ matrix W such that

$$[R1 \quad R2]V = [W \quad 0] \ , \tag{2.3}$$

W is non-singular and upper triangular.

(iv) Compute

$$\begin{array}{l} k \\ m-k \end{array} \begin{array}{l} \{ \\ \{ \end{array} \begin{bmatrix} \underline{qb1} \\ \underline{qb2} \end{bmatrix} = Q\underline{b} \tag{2.4}$$

(v) The solutions of the least squares problem

$$\begin{bmatrix} W & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \underline{y1} \\ \underline{y2} \end{bmatrix} \cong \begin{bmatrix} \underline{qb1} \\ \underline{qb2} \end{bmatrix} \tag{2.5}$$

are of the form

$$\left.\begin{array}{l} \hat{\underline{y1}} = W^{-1}\, \underline{qb1} \quad (k \text{ components}) \\ \hat{\underline{y2}} \text{ arbitrary } (n-k \text{ components}) \end{array}\right\} \tag{2.6}$$

Our subroutines set $\hat{\underline{y2}} = \underline{0}$ so that

$$\hat{\underline{y}} = \begin{bmatrix} \hat{\underline{y1}} \\ \hat{\underline{y2}} \end{bmatrix}$$

is the "minimal length solution" of (2.5).

(vi) Compute

$$\hat{\underline{x}} = PV\begin{bmatrix}\underline{y1}\\ \underline{y2}\end{bmatrix} = PV\hat{\underline{y}} \qquad (2.7)$$

This is the minimal length solution for the problem $A1\underline{x} \cong b$.

(vii) Residuals. Let $\underline{x}$ be a given vector (e.g. the minimal length solution (2.7)) and let $\underline{y} = (PV)^{-1}\underline{x}$. We may be interested in two residual vectors:

$$\underline{z1} = \underline{b} - A1\ \underline{x} \quad \text{and} \quad \underline{z2} = \underline{b} - A\ \underline{x}.$$

After the transformations described in (i)-(iv) have been performed, the matrix A (or A1) and the vector $\underline{b}$ are not available in memory anymore. They are essentially replaced by the matrix RV and by the vector $Q\underline{b}$. Therefore, for the computation of $\underline{z1}$ and $\underline{z2}$ we must use the following relations. Write

$$RV = \begin{array}{l} k\{ \\ m-k\{ \end{array}\begin{bmatrix} R1 & R2 \\ 0 & R3 \end{bmatrix}\}n-k \qquad V = \begin{bmatrix} W & 0 \\ \multicolumn{2}{c}{(R3V)} \end{bmatrix}$$

Hence,

$$RV\underline{y} = \begin{bmatrix} W\underline{y} \\ (R3V)\underline{y} \end{bmatrix}$$

and the components nr. n+1, n+2, ..., m of $RV\underline{y}$ are all equal to 0. We have

$$\underline{qb} - \begin{bmatrix} W\underline{y} \\ (R3V)\underline{y} \end{bmatrix} = \begin{cases} Q\underline{z2} \text{ if (R3V) is not ignored} & (2.8) \\ Q\underline{z1} \text{ if (R3V) is ignored (i.e. if we set R3=0)} & (2.9) \end{cases}$$

$\underline{z1}$ and $\underline{z2}$ are computed by inverting these equations.

Particularly

- If $\underline{y} = \hat{\underline{y}}$, where $\hat{\underline{y}}$ is computed as in (2.6), then the first k components of $Q\underline{z2}$ are all equal to 0 and

$$Q\underline{z1} = (\underbrace{0 \ldots 0}_{k}, \underline{qb2}^T)^T$$

- If k = n, then

$$Q\underline{z1} = Q\underline{z2} = \begin{bmatrix} \underline{qb1} - W\underline{y} \\ \underline{qb2} \end{bmatrix}\begin{array}{l} \}\ k \\ \}\ m-k \end{array}$$

(viii) The error variance in the model $A\underline{e} + \underline{error} = \underline{b}$ may be estimated by

$$\sigma^2 = \frac{1}{m-k}(\underline{z1}^T, \underline{z1}) = \frac{1}{m-k}(\underline{qb2}^T, \underline{qb2}) \qquad (2.10)$$

where $\underline{z1}$ is the residual vector corresponding to $\hat{\underline{x}}$ in (2.7).

The matrices Q and V are obtained as a product of elementary Householder transformations of the form $Q_j = I_m + b_j^{-1}\,\underline{u}^{(j)}\,\underline{u}^{(j)T}$ where the vectors $\underline{u}^{(j)}$ and the scalars b_j are conveniently chosen. The determination and the application of these transformations are obtained by the algorithm HFTI:

"For the construction of the j^{th} Householder transformation, one considers columns j through n and selects that one, say, column λ, whose sum of squares of components in rows j through m is greatest. The contents of columns j and λ are then interchanged and the j^{th} Householder transformation is constructed to zero the elements stored in a_{ij}, $i=j+1 \ldots m$."

The code of the subroutines MTRF and CLLS is very similar to the one of the subroutine HFTI in L6, p.290-291. The principal modifications are:

- The introduction of an input parameter check: these parameters must belong to a meaningful range.

- The program parts corresponding to the steps i), ii) and iii) are grouped in MTRF and the others in CLLS.

- Residuals are computed following vii).

- The simultaneous handling of several vectors $\underline{b}$ is eliminated. The memory organization of the matrices A, R, W, Q, V, P (and of some auxiliary arrays: G, H, IP) is the same as it is described in L6, p.80.

3. ALGORITHM FOR THE COMPUTATION OF THE ROBUST SOLUTION

We consider now the subroutine HREG to compute the solution of Problem B. HREG implements a modification of the algorithm H of L4 in the special case of a linear problem. The modification consists of the introduction of the weights w_i-s and of some minor changes due to the use of the relations (2.1)-(2.7). We will discuss at the end of this section the reason of the introduction of the w_i-s and how they should be chosen.

We assume that

- $\underline{x}$ and σ initially have some starting values (these may be given by the minimal length solution of $A\underline{x} \cong \underline{b}$) .

- The transformations Q, P and V are previously computed (and applied to A whose pseudorank is k).

- ε is a given tolerance and MAXIT a given integer.

- The unscaled covariance matrix (COV) of the estimate $\underline{y}$ (see below, steps 7 and 8) is previously computed (using UCVY) and given to HREG as a parameter.

HREG performs the following steps:

- Computation of $\underline{y} = (PV)^{-1}\underline{x}$

- Computation of $\underline{qb} = Q\underline{b}$ (only if $\underline{qb}$ was not previously computed)

- Computation of BETA = $\int(\phi^2(x)/2)\exp(-x^2/2)/\sqrt{2\pi}\,dx$ with $\phi(x) = \max(-c, \min(c,x))$ (c a given parameter), unless the user prefers a different function ϕ or a different value of BETA. In this case the new ϕ has to be externally defined together with the new BETA. (See also the remark 3 at the end of this section.)

- Computation of

$$\mathrm{CONST} = (m-k)\ \mathrm{BETA}.$$

Then an iterative algorithm is started:

<u>Step 1</u>. Set NIT = 1

<u>Step 2</u>. Compute

$$\underline{qz1} = \underline{qb} - RV\underline{y}$$

and $\underline{z1} = Q^{-1}\,\underline{qz1}$ following (2.9)

<u>Step 3</u>. Compute a new value for σ from:

$$s^2 := \frac{1}{\mathrm{CONST}} \sum_{i=1}^{m} \chi\left(\frac{z1_i}{w_i\sigma}\right)(w_i\sigma)^2$$

and $\sigma := \sqrt{s^2}$

<u>Step 4</u>. "Winsorise" the residuals

$$z1_i = \psi\left(\frac{z1_i}{w_i\sigma}\right)(w_i\sigma)$$

<u>Steps 5 and 6</u>. Compute

$\underline{qz1} := Q\underline{z1}$ and the vector $\underline{zy}$ such that

$$\begin{bmatrix} zy_1 \\ \vdots \\ zy_k \end{bmatrix} = W^{-1} \begin{bmatrix} qz1_1 \\ \vdots \\ qz1_k \end{bmatrix}$$

$$\begin{bmatrix} zy_{k+1} \\ \vdots \\ zy_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

<u>Step 7</u>. Set $\underline{y} = \underline{y} + \underline{zy}$

<u>Step 8</u>. Stop iterations and go to 9 if the (transformed) parameter estimates change by less than ε times their standard deviation, i.e. if for $j = 1 \ldots k$

$$|zy_j| < \varepsilon\sqrt{(\mathrm{COV})_{jj}}\,\sigma$$

and the absolute change of σ is less than ε or if $\mathrm{NIT} \geq \mathrm{MAXIT}$; otherwise put NIT: = NIT + 1 and go to Step 2.

<u>Step 9</u>. Compute

$$\underline{x} = PK\,\underline{y}$$

$\underline{z1}$ and $\underline{z2}$ following (2.8) and (2.9)

$\underline{zx} := PK\,\underline{zy}$ (= last parameter change in the original coordinate system).

<u>Remark 1</u>. Optimal choice of the weights w_i-s.

We now discuss briefly how the w_i-s have to be chosen in order to get an estimator which is optimal for an Hampel-extremal problem. See L1, L2, L5 and L8 for further information. We consider the linear model

$$A\underline{\theta} + \underline{r} = \underline{b}$$

where $\underline{\theta} \in \mathbb{R}^n$ is a vector of parameters,

A is a matrix of random row vectors in $\mathbb{R}^n$ with a given distribution (we denote the i^{th} row of A by $\underline{a}_i = [a_{i1} \ldots a_{in}])$.

$\underline{b} \in \mathbb{R}^m$ is the (so called) "observation vector",

$\underline{r} \in \mathbb{R}^m$ is the "error vector", and

$$\begin{cases} \underline{r} \sim N(\underline{0}, \sigma^2 I_m), \ \sigma \text{ is known} \\ \underline{r} \text{ is independent of } A \end{cases} \qquad (4.1)$$

M-estimators of $\underline{\theta}$ are given by a family of functions

$$\{\underline{\psi}(b,\underline{a},\underline{\theta}),\ b\in\mathbb{R},\ \underline{a}\in\mathbb{R}^n,\ \underline{\theta}\in\mathbb{R}^n\}$$

into the $\mathbb{R}^n$ which define the estimate $\underline{\tilde{\theta}}$ as a selected solution of

$$\int \underline{\psi}(b,\underline{a},\underline{\theta})\, dP_n(b,\underline{a}) = \underline{0} \qquad (4.2)$$

where P_n is the empirical measure defined by the "observations" $(b_i,\underline{a}_i)$. Fisher consistency means

$$\int \underline{\psi}(b,\underline{a},\underline{\theta})\, dP_{\underline{\theta}}(b,\underline{a}) \equiv \underline{0} \quad \text{(for all } \theta) \qquad (4.3)$$

where $P_{\underline{\theta}}(b,\underline{a})$ is the family of distributions given by (4.1).

We will now restrict our attention to the family of Fisher consistent M-estimators defined by the functions $\underline{\psi}$ *of the form*

$$\underline{\psi}(b,\underline{a},\underline{\theta}) = \psi\left(\frac{b-a\theta}{\sigma w(\underline{a})}\right) w(\underline{a})\underline{a}$$

where ψ is a function of $\mathbb{R}$ into $\mathbb{R}$ and w a function of $\mathbb{R}^n$ into $\mathbb{R}$.

In this class the influence curve of an estimator $\underline{\tilde{\theta}}$ is

$$IC(b,\underline{a};\underline{\tilde{\theta}}) = E\left[\underline{a}^T\psi'\left(\frac{r}{\sigma w(\underline{a})}\right)\underline{a}\right]^{-1} \sigma\underline{a}^T w(\underline{a})\psi\left(\frac{r}{\sigma w(\underline{a})}\right)$$

where E denotes expectation taken for the measure $P_{\underline{\theta}}$ and $\psi'(x) = \frac{d}{dx}\psi(x)$. Our aim is to minimize the trace of the asymptotic covariance matrix $E(\underline{IC}\cdot\underline{IC}^T)$ of $\underline{\tilde{\theta}}$, under the condition $\underline{IC} \le \sigma\underline{c}$ where $\underline{c}$ is an n-vector whose components are all equal to a given number c. *If we put* $\underline{a}'^T = H\underline{a}^T$, where H is an $n\times n$ symmetric and non-singular matrix, we get

$$\underline{IC}(b,\underline{a};\underline{\tilde{\theta}}) = H^T E\left[\underline{a}'^T\psi'\left(\frac{r}{\sigma w'(\underline{a}')}\right)\underline{a}'\right]^{-1} \sigma\underline{a}'^T w'(\underline{a}')\psi\left(\frac{r}{\sigma w'(\underline{a}')}\right) \qquad (4.5)$$

where $w'(\underline{a}') = w\left[\underline{a}'(H^T)^{-1}\right] = w(\underline{a})$.

We can now choose H so that

$$H^T E\left[\underline{a}'^T\psi'\left(\frac{r}{\sigma w'(\underline{a}')}\right)\underline{a}'\right]^{-1} = I_n \qquad (4.6)$$

and then it is easy to see that the solution of the extremal problem is given by

$$\psi(x) = \psi_c(x) = \begin{cases} -c & x \le -c \\ x & -c < x < c \\ c & c \le x \end{cases} \qquad (4.7)$$

and

$$w'(\underline{a}') = 1/\left[\underline{a}'\underline{a}'^T\right]^{1/2}. \qquad (4.8)$$

If the expectation in (4.6) is taken for the product of the empirical distribution of $\underline{a}_1,\ldots,\underline{a}_m$ with the multinormal distribution of $\underline{r}$, we obtain that the weights $w_i = w(\underline{a}_i)$ have to be determined by solving for H:

$$\begin{cases} HA^T\,\mathrm{Diag}\left[E_\phi\left(\psi_c'\left(\frac{r}{\sigma w_i}\right)\right)\right]AH^T = H \\ w_i = 1 / \left(\underline{a}_i H^T H \underline{a}_i^T\right)^{1/2} \end{cases} \qquad (4.9)$$

where $\mathrm{Diag}(\delta_i)$ denotes a diagonal matrix with elements $\delta_1 \ldots \delta_n$ and

$$E_\phi\left(\psi_c'\left(\frac{r}{\sigma w_i}\right)\right) = 2\phi(w_i c) - 1 \qquad (4.10)$$

with $$\phi(x) = \int_{-\infty}^{x} \exp(-r^2/2) / \sqrt{2\pi}\, dr . \tag{4.11}$$

The system (4.9) has to be solved numerically.

For a slightly different extremal problem this system is replaced by a simpler one. Let $\underline{\theta}^* = (H^T)^{-1}\underline{\theta}$ where H is chosen so that

$$E\left[\underline{a}'^T \psi\cdot\left(\frac{r}{\sigma w'(\underline{a}')}\right)\underline{a}'\right] = I_n . \tag{4.12}$$

Then the influence function of an "M-estimator" $\underline{\tilde{\theta}}^* = (H^T)^{-1}\underline{\tilde{\theta}}$ of $\underline{\theta}^*$ is

$$\underline{IC}^*(b,\underline{a};\underline{\tilde{\theta}}^*) = \sigma\underline{a}'^T w'(\underline{a}')\psi\left(\frac{r}{\sigma w'(\underline{a}')}\right) . \tag{4.13}$$

In order to minimize the trace of $E(\underline{IC}^*\cdot\underline{IC}^{*T})$ we have to choose ψ and $w'(\underline{a}')$ as in (4.7) - (4.8) and (4.9) becomes

$$A'^T \operatorname{Diag}\left(2\phi\left(\frac{c}{\sqrt{\underline{a}'\underline{a}'^T}}\right) - 1\right)A' = I_n . \tag{4.14}$$

(4.14) has exactly the form of equation (15.3), p. 41 in L4 by which affine invariant M-estimators of the covariance matrix of $\underline{a}$ are defined. A program for their computation may then be used for the numerical determination of the w_i-s.

<u>Remark 2</u>.

By combining the main subroutines in different ways we can get several different computational schemes. For example,

- Given starting values $\underline{x}^0$ and σ^0 for $\underline{x}$ and σ, iterate fully until the tolerance ε in Step 8 is reached.

- Given starting values $\underline{x}^0$ and σ^0, compute a certain number of iterations using a monotone ψ and a few ones with a redescending ψ function (see L3 for a motivation).
- As we assume that the triangularization of A is computed separately before the call of HREG, it is possible to solve problems economically with the same design matrix A but different right sides $\underline{b}$.

With our tools, the simplest starting values, computationally, for $\underline{x}^0$ and σ^0, are those produced by CLLS, despite their poor robustness properties. ("Most customers will want to see the least squares result anyway!", see L4.). Some of the possible computational schemes are given as examples in the subroutine RLLS.

<u>Remark 3</u>

When the w_i-s are all equal to 1, the value of σ, computed in step 3 of HREG, is a robust estimate of the error standard deviation which is asymptotically unbiased for normal errors. This value is used to standardize the residuals. If $(m-k)/m$ is not close to 1, it seems better to standardize using $\sqrt{(m-k)/m}\,\sigma$ instead of σ. This may be easily done as follows:

- Put the starting value of σ (i.e. SIGMA in the parameter list of HREG) equal to an estimate of the residual standard error.
- Use externally defined BETA and ϕ such that

$$BETA = \frac{m}{m-k} E(\chi) := \frac{m}{m-k} \int \chi(x)\exp(-x^2/2)/\sqrt{2\pi} \ dx .$$

In this way the quantities computed in step 3 are

$$s^2 := \frac{1}{mE(\chi)} \sum_{i=1}^{m} \chi\left(\frac{z1_i}{\sigma}\right) \sigma^2$$

and $\sigma := \sqrt{s^2}$.

(Remember to multiply the final value of σ by $\sqrt{m/(m-k)}$ for estimating the error standard deviation).

Another possible standardization of the residuals uses weights $w_i = \sqrt{1-ht_i}$, where ht_i, $i = 1 \ldots m$, are the diagonal elements of the Hat matrix

$$(HT) = A(A^TA)^{-1}A^T .$$

The code of a FORTRAN subroutine for the computation of the ht_i - s is given in Appendix 2.

When the w_i-s are not all equal to 1, it may be appropriate to use an eternally defined BETA such that

$$BETA = \int \chi(x/w(x))\, w(x)^2 \exp(-x^2/2)/\sqrt{2\pi}\ dx$$

with an appropriate function w .

4. COVARIANCE MATRIX OF THE PARAMETER ESTIMATES

In this section we restrict our discussion to problems with all the w_i-s in (1.1) - (1.2) equal to 1. Correspondent subroutines for the more general case are still to be written.

A. Unscaled covariance matrix of the parameter estimates in a problem with upper triangular or upper trapezoidal design matrix.

We have seen that the unscaled covariance matrix (COV) of the estimate $\underline{y}$, defined at step 7 of HREG, has to be computed before calling HREG. In the case $w_i = 1$ for all i, this matrix coincides with the unscaled covariance matrix of the least squares estimate $\hat{\underline{y}}$ defined in Section 2, (v). As $\underline{\hat{y}2}$ is set equal to 0 by CLLS, we define the unscaled covariance matrix of $\hat{\underline{y}}$ as follows:

$$(\mathrm{COV}) = \left.\begin{bmatrix} (W^T W)^{-1} & \vdots & 0 \\ \cdots\cdots & & \cdots\cdots \\ \underbrace{0}_{k} & \vdots & \underbrace{0}_{n-k} \end{bmatrix}\right. \begin{matrix} \}\,k \\ \\ \}\,n-k \end{matrix} .$$

The subroutine UCVY computes (COV). Its code is taken from the algorithm COV in L6, p. 69.

B. Scaled covariance matrix of the parameter estimates

The parameter estimates in the original coordinate system are $\hat{\underline{x}} = PV\hat{\underline{y}}$ (see (2.7)) or $\underline{x} = PV\underline{y}$ (see step 9 in HREG). Hence their unscaled covariance matrix is

$$(\mathrm{COVX}) = PV(\mathrm{COV})V^T P^T \quad .$$

The subroutine SCVX computes this matrix given the matrices (COV), P and V and multiplies it by a scale factor. This may be given as a parameter (e.g., using the value of σ computed by CLLS from (2.10));

otherwise, it is computed as

$$\kappa^2 \frac{1}{m-k} \Sigma \psi \left(\frac{z1_i}{\sigma}\right)^2$$

where $\underline{z1}$ is the residual vector computed in step 9 of HREG. κ is a correction factor (see L4, p. 40):

$$\kappa = (1+(k/m))\,\mathrm{var}\ \psi' / (\mathrm{ave}\ \psi')^2 / (\mathrm{ave}\ \psi')$$

$$\mathrm{ave}\ \psi' = \frac{1}{m} \Sigma \psi' \left(\frac{z1_i}{\sigma}\right)$$

$$\mathrm{var}\ \psi' = \frac{1}{m} \Sigma \left(\psi' \left(\frac{z1_i}{\sigma}\right) - \mathrm{ave}\ \psi'\right)^2 \quad .$$

σ is computed in step 3 of HREG.

APPENDIX 1. **Guide for the Auxiliary and Utility Subroutines**

Auxiliary:

RES : computes residuals following (vii) in Section 2.

R3V : applies Householder transformations with the special purpose to compute the matrix (R3V) of (vii), Section 2.

SOLVE : solves a triangular system of linear equations (back substitution) (see §2.6 and Steps 5-6 in HREG).

PVMIX : given $\underline{x}$, this subroutine computes $\underline{x} = (PV)^{-1}\underline{x}$ (see preliminary steps in HREG).

NEWSIG : computes a new value for the robust estimate of the error standard deviation (see Step 3 in HREG).

HUB : computes "Winsorised" residuals (see Step 4 in HREG).

FACS : computes the robust scale factor for the estimate co-variance matrix.

ADIAG : exchanges the contents of two arrays.

DIFF (Function) : computes a difference (see L6, p.278).

PSI (Function) : computes the Huber function

$$\psi_c(x) = \max(-c, \min(c,x)) .$$

CHI (Function) : computes $\chi(x) = \psi_c(x)^2/2$.

PSIPRM (Function): computes the first derivative of $\psi_c(x)$.

PSIU (Function) : computes the function $\psi(x)$ (see introduction) and has to be completed by the user.

CHIU (Function) : computes $\chi(x)$ (see introduction) and has to be completed by the user.

PSIPRU (Function): computes the function $\frac{d}{dx}\psi(x)$ and has to be completed by the user.

Utility :

H12 : construction and/or application of a single elementary Householder transformation (see L6, p.308).

EXCH : exchanges two rows and two columns of a symmetric matrix.

VSV : computes the matrix VSV, where V is an elementary Householder transformation and S a symmetric matrix stored columnwise in a one-dimensional array.

PHI (Function) : cumulative normal distribution.

APPENDIX 2. **FORTRAN Subroutine for the Computation of the Diagonal Elements of the Hat Matrix.**

This subroutine uses the output (arrays A and H) of MTRF and the subroutine H12 to compute the diagonal elements HT(1) ... HT(m) of the Hat matrix

$$(HT) = A(A^TA)^{-1}A^T .$$

It is supposed that the pseudorank of A is m. SC is an auxiliary array. The integer parameters M, N, MDA have the same meaning as in MTRF.

Method: From $A = Q^TRP^T$ we get

$$(HT) = Q^T \left[\begin{array}{c|c} I_n & 0 \\ \hline 0 & 0 \end{array}\right] Q \quad \begin{array}{l} \}n \\ \}(m-n) \end{array}$$

Computer Listing:

```
      SUBROUTINE HAT (A,HT,H,SC,M,N,MDA)
      DIMENSION A(MDA,N),HT(M),H(N),SC(M)
      DOUBLE PRECISION SM,DZERO
      DZERO = 0.DO
      DO 100 I = 1,M
      DO 20 J = 1,M
   20 SC(J) = 0.
      SC(I) = 1.
      DO 50 JJ = 1,N
      J = JJ
   50 CALL H12(2,J,J+1,M,A(1,J),1,H(J),SC,1,M,1,M)
      SM = DZERO
      DO 70 J = 1,M
   70 SM = SM+SC(J)*DBLE(SC(J))
  100 HT(I) = SM
      RETURN
      END
```

REFERENCES

L1. F.R. Hampel (1968), Contributions to the theory of robust estimation, Ph.D Thesis, University of California Berkeley.

L2. F.R. Hampel(1978a), Optimally Bounding the Gross-error-sensitivity and the influence of position in factor space, (Invited paper read at the ASA/IMS Meeting in San Diego, Ca., August 16, 1978); Research report No. 18, Fachgruppe fuer Statistik, ETH Zuerich.

L3. F.R. Hampel(1978b), Modern trends in the theory of robustness, <u>Math. Operationsforsh. Statist.</u>, <u>Ser. Statistics</u>, <u>Vol. 9</u>, No. 3, p.425-442.

L4. P.J. Huber (1977) , <u>Robust Statistical Procedures</u>, SIAM Regional Conference series in applied mathematics, No. 27.

L5. W.S. Krasker(1977), Parametric estimation in approximate parametric models. Unpubl. manuscript, Univ. of Michigan.

L6. C.L. Lawson and R.J. Hanson (1974), <u>Solving Least Squares Problems</u>, Prentice Hall, Englewood Cliffs, N.J.

L7. W. Stuetzle and T. Gasser, Richtlinien fuer die Programmierung von Subroutinen, Fachgruppe fuer Statistik, ETH, Zuerich.

L8. W. Stahel (Oct. 1978), personal note.